Module 9 Lab activity

Neil Yetz & Gemma Wallace

Contents

load libraries

```
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
## Warning: package 'tibble' was built under R version 3.6.3
```

```
## Warning: package 'tidyr' was built under R version 3.6.3
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
## Warning: package 'forcats' was built under R version 3.6.3
```

```
library(psych)
library(olsrr)
```

```
## Warning: package 'olsrr' was built under R version 3.6.3
```

```
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 3.6.3
```

Import data

```
bac <- read_csv("bac_obs.csv")
```

```
## Parsed with column specification:
## cols(
##   id = col_double(),
##   weight = col_double(),
##   typ_drks = col_double(),
##   alcexp = col_double(),
##   pmood = col_double(),
##   absorb = col_double(),
##   alc_gm = col_double(),
##   bac = col_double()
## )
```

Describe the data

```
describe(bac)
```

```
##          vars   n   mean    sd median trimmed   mad   min    max  range  skew
## id          1 200 100.50 57.88 100.50  100.50 74.13  1.00 200.00 199.00  0.00
## weight      2 200  68.46  9.92  68.90   68.52  9.56 37.60  91.40  53.80 -0.18
## typ_drks    3 200  46.86 14.67  46.50   47.11 14.08  4.00  79.00  75.00 -0.15
## alcexp      4 200   4.09  0.78   4.13    4.11  0.77  2.01   6.06   4.05 -0.26
## pmood       5 200   5.12  1.40   5.00    5.09  1.48  1.00   9.00   8.00  0.18
## absorb      6 200   4.69  0.91   4.64    4.68  0.95  2.67   6.80   4.13  0.11
## alc_gm      7 200  32.79  7.73  33.00   32.76  8.90  8.00  58.00  50.00  0.04
## bac         8 200   0.08  0.02   0.08    0.08  0.02  0.02   0.15   0.13  0.22
##          kurtosis   se
## id          -1.22 4.09
## weight       0.16 0.70
## typ_drks    -0.29 1.04
## alcexp      -0.07 0.06
## pmood       -0.05 0.10
## absorb      -0.60 0.06
## alc_gm       0.43 0.55
## bac          0.40 0.00
```

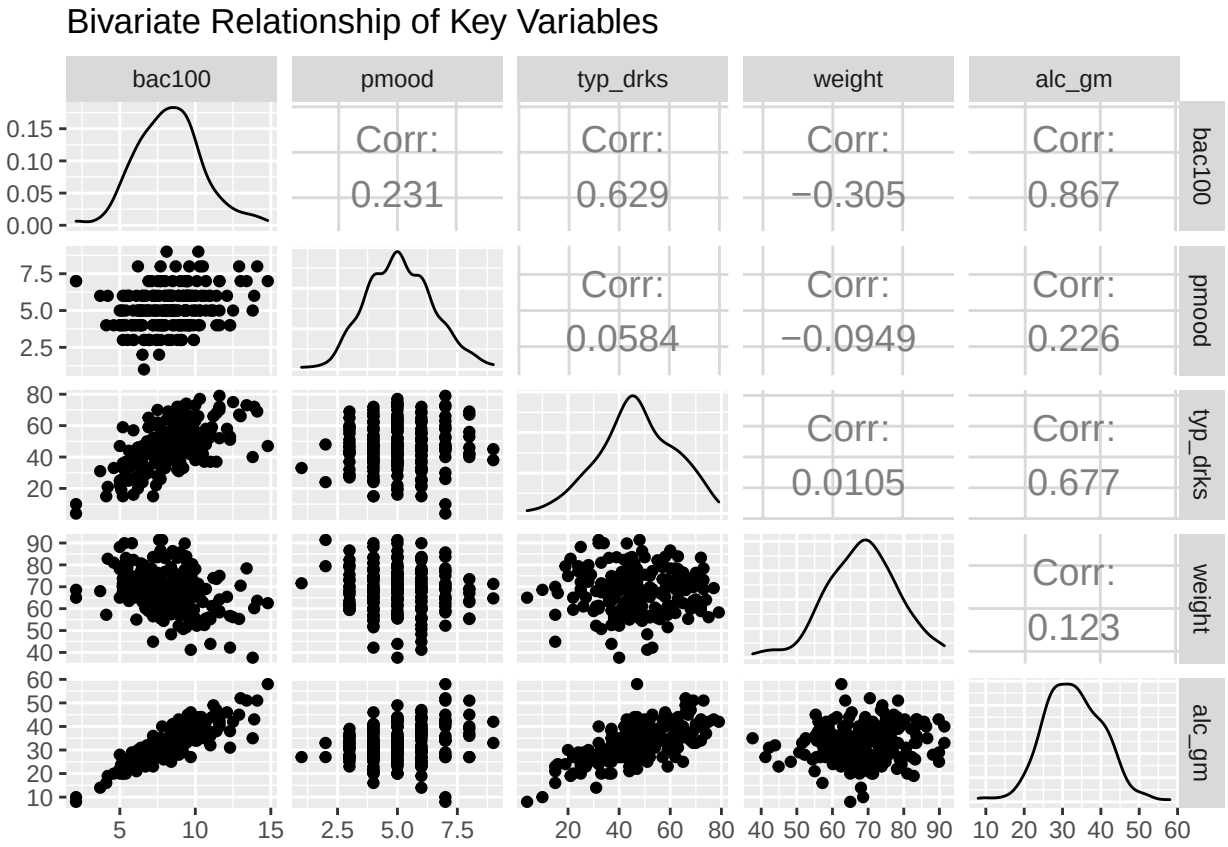
Mutate BAC variable

```
bac <- mutate(bac, bac100 = bac*100)
```

Get Sccatterplot matrix

```
scatterplot2 <- ggpairs(bac, columns = c("bac100", "pmood", "typ_drks", "weight", "alc_gm"),
upper = list(continuous = wrap("cor", size=5)),
title = "Bivariate Relationship of Key Variables", progress = ggmatrix_progress(clear = TRUE))

print(scatterplot2)
```



Fit Regression

Regress BAC100 on weight

```
m1 <- lm(data = bac, bac100 ~ weight)
ols_regress(m1)
```

```
##                         Model Summary
## ----------------------------------------------------------------
## R                       0.305       RMSE                 2.065
## R-Squared               0.093       Coef. Var           24.924
## Adj. R-Squared          0.088       MSE                  4.265
## Pred R-Squared          0.075       MAE                  1.638
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                Sum of
##               Squares         DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression     86.374          1         86.374    20.253    0.0000
## Residual      844.414        198          4.265
```

```
## Total         930.788        199
## ----------------------------------------------------------------
##
##                                 Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ----------------------------------------------------------------------------------------
## (Intercept)    12.832         1.021                12.571    0.000    10.819    14.845
##      weight    -0.066         0.015       -0.305    -4.500    0.000    -0.096    -0.037
## ----------------------------------------------------------------------------------------
```

Regress BAC100 on alc_gm

```
m2 <- lm(data = bac, bac100 ~ alc_gm)
ols_regress(m2)
```

```
##                         Model Summary
## ---------------------------------------------------------------
## R                        0.867       RMSE                 1.082
## R-Squared                0.751       Coef. Var           13.056
## Adj. R-Squared           0.750       MSE                  1.170
## Pred R-Squared           0.747       MAE                  0.799
## ---------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                Sum of
##               Squares         DF    Mean Square       F         Sig.
## --------------------------------------------------------------------
## Regression    699.106          1        699.106     597.47    0.0000
## Residual      231.682        198          1.170
## Total         930.788        199
## --------------------------------------------------------------------
##
##                                 Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower    upper
## ----------------------------------------------------------------------------------------
## (Intercept)    0.332         0.334                 0.994    0.321    -0.327    0.991
##      alc_gm    0.243         0.010        0.867    24.443    0.000     0.223    0.262
## ----------------------------------------------------------------------------------------
```

Regress BAC100 on weight and alc_gm

```
m3 <- lm(data = bac, bac100 ~ alc_gm + weight)
ols_regress(m3)
```

```
##                         Model Summary
## ---------------------------------------------------------------
## R                        0.961       RMSE                 0.604
## R-Squared                0.923       Coef. Var            7.284
```

```
## Adj. R-Squared          0.922       MSE                 0.364
## Pred R-Squared          0.920       MAE                 0.449
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------------
##                Sum of
##               Squares         DF    Mean Square       F          Sig.
## --------------------------------------------------------------------------
## Regression    859.029          2        429.514    1179.142    0.0000
## Residual       71.759        197          0.364
## Total         930.788        199
## --------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------
## (Intercept)     6.094         0.332                 18.341    0.000     5.439     6.750
##      alc_gm     0.257         0.006        0.918    46.056    0.000     0.246     0.268
##      weight    -0.091         0.004       -0.418   -20.953    0.000    -0.100    -0.082
## ----------------------------------------------------------------------------------------------
```

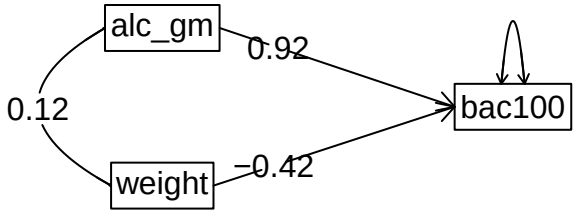
Subset data for correlation matrix

```
bac_subset <- select(bac, bac100, alc_gm, weight)
cor_matrix <- cor(bac_subset)
```

Calculate regression from correlation matrix

```
set.cor(y = ("bac100"), x = c("alc_gm", "weight"), z = NULL, data = cor_matrix)
```

Regression Models




```
## Call: setCor(y = y, x = x, data = data, z = z, n.obs = n.obs, use = use,
##     std = std, square = square, main = main, plot = plot, show = show)
##
## Multiple Regression from matrix input
##
##  DV =  bac100
##        slope  VIF
## alc_gm  0.92 1.02
## weight -0.42 1.02
##
##  Multiple Regression
##           R   R2  Ruw R2uw
## bac100 0.96 0.92 0.88 0.78
```